

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) NOVEMBER 2015		2. REPORT TYPE JOURNAL ARTICLE (POST PRINT)		3. DATES COVERED (From - To) JUN 2013 – AUG 2015	
4. TITLE AND SUBTITLE IDENTIFYING CODES ON DIRECTED DE BRUIJN GRAPHS				5a. CONTRACT NUMBER NA	
				5b. GRANT NUMBER NA	
				5c. PROGRAM ELEMENT NUMBER 62788F	
6. AUTHOR(S) Victoria Horan, Debra Boutin, Mikko Pelto				5d. PROJECT NUMBER HORN	
				5e. TASK NUMBER IN	
				5f. WORK UNIT NUMBER HO	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/Information Directorate Rome Research Site/RITF 525 Brooks Road Rome NY 13441-4505				8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/Information Directorate Rome Research Site/RITF 525 Brooks Road Rome NY 13441-4505				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RI	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-RI-RS-TP-2015-004	
12. DISTRIBUTION AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA Case Number: 88ABW-2015-3797 DATE CLEARED: 4 AUG 15					
13. SUPPLEMENTARY NOTES This work may be copyrighted. One or more of the authors is a U.S. Government employee working within the scope of their Government job; therefore, the U.S. Government is joint owner of the work and has the right to copy, distribute, and use the work. All other rights are reserved by the copyright owner.					
14. ABSTRACT For a directed graph G , a t -identifying code is a subset $S \subseteq V(G)$ with the property that for each vertex $v \in V(G)$ the set of vertices of S reachable from v by a directed path of length at most t is both non-empty and unique. A graph is called t -identifiable if there exists a t -identifying code. This paper shows that the directed de Bruijn graph $B(d, n)$ is t -identifiable for $n \geq 2t-1$, and is not t -identifiable for $n \leq 2t-2$.					
15. SUBJECT TERMS Identifying Code; De Bruijn Network; Graph Theory					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 27	19a. NAME OF RESPONSIBLE PERSON VICTORIA HORAN
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) N/A

Identifying Codes on Directed De Bruijn Graphs

Debra Boutin *

Department of Mathematics
Hamilton College

Victoria Horan †

Air Force Research Laboratory
Information Directorate

Mikko Pelto ‡

Department of Mathematics and Statistics
University of Turku

August 27, 2015

Abstract

For a directed graph G , a t -*identifying code* is a subset $S \subseteq V(G)$ with the property that for each vertex $v \in V(G)$ the set of vertices of S reachable from v by a directed path of length at most t is both non-empty and unique. A graph is called t -*identifiable* if there exists a t -identifying code. This paper shows that the de Bruijn graph $\vec{B}(d, n)$ is t -identifiable for $n \geq 2t - 1$, and is not t -identifiable for $n \leq 2t - 2$. This

*dboutin@hamilton.edu

†victoria.horan.1@us.af.mil

‡mikko.pelto@utu.fi, research supported by the Finnish Cultural Foundation

Approved for public release; distribution unlimited: 88ABW-2015-3797.

paper also proves that a t -identifying code for t -identifiable de Bruijn graphs must contain at least $d^{n-1}(d-1)$ vertices. Constructions are given to show that this lower bound is achievable for t -identifying codes when $n \geq 2t$. Further a (possibly) non-optimal construction is given when $n = 2t - 1$. Additionally, with respect to $\vec{\mathcal{B}}(d, n)$ this paper proves upper and lower bounds on the size of a minimum t -dominating set (a subset with the property that every vertex is at distance at most t from the subset), that the minimum size of a *directed resolving set* (a subset with the property that every vertex of the graph can be distinguished by its directed distances to vertices of S) is $d^{n-1}(d-1)$, and that if $d > n$ the minimum size of a *determining set* (a subset S with the property that the only automorphism that fixes S pointwise is the trivial automorphism) is $\lceil \frac{d-1}{n} \rceil$.

1 Introduction

First introduced in 1998 [9], an *identifying code* for a graph G is a subset $S \subseteq V(G)$ with the property that for each $v \in V(G)$ the subset of vertices of S that are adjacent to v is non-empty and unique. That is, each vertex of the graph is uniquely identifiable by the non-empty subset of vertices of S to which it is adjacent. Note that not all graphs have an identifying code; those that do are called *identifiable*. A graph fails to be identifiable if and only if it contains a pair of vertices with the same closed (in-)neighborhood; such vertices are called *twins*. Extending these definitions to t -identifying and t -twins is easy and is covered in Section 3. Identifying codes can be quite useful in applications. For example, an identifying code in a network of smoke detectors allows us to determine the exact location of a fire given only the set of detectors that have been triggered. However, the problem of finding identifying codes is NP-Hard [5]. The computational cost has so far limited the real-world use of identifying codes.

The directed de Bruijn graph $\vec{\mathcal{B}}(d, n)$ is a directed graph in which the vertices are strings of length n from an alphabet \mathcal{A} with d letters, and with a directed arc from vertex $x_1x_2 \dots x_n$ to vertex $x_2 \dots x_na$ for each $a \in \mathcal{A}$. When looking for a graph model for applications, it is useful to choose a graph with relatively few edges, but many short paths between any pair of vertices [2]. The de Bruijn graphs have both of these desirable properties. In addition, given an arbitrary pair of vertices in a de Bruijn graph, there

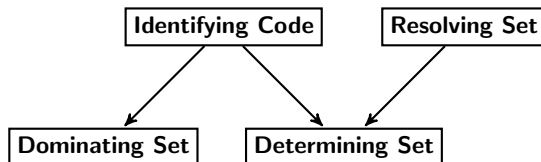


Figure 1: Vertex Subset Relationships

are routing algorithms that, with high probability, create a path of length $O(\log n)$ between the pair [13]. The properties of de Bruijn graphs enable some problems that are NP-complete on general graphs, such as the Hamilton cycle problem, to be computationally solvable on de Bruijn graphs [14]. We will see that for most directed de Bruijn graphs, the construction of minimum t -identifying codes is indeed solvable (when they exist).

Other vertex subsets that we consider for de Bruijn graphs in this paper are dominating sets, resolving sets, and determining sets. Dominating sets provide complete coverage for a graph, while resolving sets provide an identification of vertices in graphs using relative distances. Finally, determining sets provide a set of vertices that is only fixed pointwise by the trivial automorphism. These types of subsets are also useful in applications. For example, resolving sets have been used in aiding the navigation of robots when distances to sufficient landmarks are known [10], and determining sets are useful in graph distinguishing which can reduce graph symmetry to enhance recognition. These different vertex subsets are interrelated, as shown in Figure 1. For example, each resolving set and identifying code is also a determining set, but not vice-versa. This is discussed more fully in Section 4.1 and Section 4.3.

In Section 2, we give careful definitions necessary for working with directed de Bruijn graphs. In Section 3, we prove that for every t -identifiable de Bruijn graph, any t -identifying code must contain at least $d^{n-1}(d-1)$ vertices. We prove by construction that this bound is achievable for t -identifying codes when $n \geq 2t$. For $n = 2t - 1$ we show the existence of a t -identifying code of size $d^{n-1}(d-1) + d^t$, which may or may not be optimal size. Furthermore, when $n \leq 2t - 2$, we are able to show that $\vec{\mathcal{B}}(d, n)$ has no t -identifying code due to the existence of t -twins. In Section 4 we study dominating sets, directed resolving sets, and determining sets for $\vec{\mathcal{B}}(d, n)$. Section 4.2 gives a proof that the minimum size of a directed resolving set for $\vec{\mathcal{B}}(d, n)$ is

$d^{n-1}(d-1)$, and Section 4.3 that the minimum size of a determining set is $\lceil \frac{d-1}{n} \rceil$.

2 Definitions

We will be considering various types of vertex subsets on the class of directed de Bruijn graphs. The following definitions will be useful in working with this class of graphs. We will use the notation $[x] = \{1, 2, \dots, x\}$.

Definition 2.1. Let $\mathcal{A}_d = \{0, 1, \dots, d-1\}$ and let \mathcal{A}_d^n be the set of all strings of length n made up of **letters** of \mathcal{A} . When d is clear from context we will use \mathcal{A} and \mathcal{A}^n respectively.

Definition 2.2. The **directed de Bruijn graph**, denoted $\vec{\mathcal{B}}(d, n)$, has vertex set \mathcal{A}_d^n . An edge from vertex $x_1x_2 \dots x_n$ to vertex $y_1y_2 \dots y_n$ exists if and only if $x_2x_3 \dots x_n = y_1y_2 \dots y_{n-1}$.

Definition 2.3. The **concatenation of two strings** $x = x_1x_2 \dots x_i$ and $y = y_1y_2 \dots y_k$ is given by $x \oplus y = x_1x_2 \dots x_iy_1y_2 \dots y_k$.

Definition 2.4. The **concatenation of sets of strings** S and T is given by $S \oplus T = \{x \oplus y \mid x \in S \text{ and } y \in T\}$.

Definition 2.5. The **prefix** of a string $x = x_1x_2 \dots x_n$ is the substring $x_1x_2 \dots x_{n-1}$, denoted by x^- .

Definition 2.6. The **suffix** of a string $x = x_1x_2 \dots x_n$ is the substring $x_2x_3 \dots x_n$, denoted by x^+ .

Definition 2.7. When discussing substrings of a string $x_1x_2 \dots x_n$, we will use the notation $x(a : b)$ to denote the substring $x_ax_{a+1} \dots x_b$.

Definition 2.8. If a string $x = x_1x_2 \dots x_n$ contains a constant substring $x(a, b) = zz \dots z$, then we will denote the consecutive letters as z^{b-a} , the constant raised to the power denoting length. This will also be used for repeated substrings, such as $0101 \dots 01 = (01)^k$.

Definition 2.9. Let $w = w_1 \dots w_n \in \mathcal{A}_d^n$. Define $w^{(t,m)} = w_1^{(t,m)} \dots w_n^{(t,m)}$ such that:

$$w_i^{(t,m)} = \begin{cases} w_t + m \pmod{d}, & \text{if } i = t; \\ w_i, & \text{otherwise.} \end{cases}$$

Definition 2.10. Let $w = w_1 \dots w_n \in \mathcal{A}_d^n$ and $\ell \in \mathbb{Z}^+$ such that $n \geq 2\ell$. Then we say that w has *period length* ℓ if $w_i = w_{i+\ell}$ for all $i \in [n - \ell]$. If we have $n < 2\ell$, then we say that w has *almost period length* ℓ .

Definition 2.11. Let $w \in \mathcal{A}_d^n$, and suppose that w has period length ℓ , and does not have period length k for any $k < \ell$. Then w is called *ℓ -periodic*.

Definition 2.12. Let $w \in \mathcal{A}_d^n$. If there exists some $\ell > \frac{n}{2}$ and word $w' \in \mathcal{A}_d^{2\ell-n}$ such that $w \oplus w'$ is ℓ -periodic, then w is called *almost ℓ -periodic*.

We now provide some lemmas regarding string properties that will be used later in the paper.

Lemma 2.13. [7] Let $\ell_1 > \ell_2$ and w be a word of length $n \geq \ell_1 + \ell_2 - \gcd(\ell_1, \ell_2)$. If w has periods (or almost periods) of length ℓ_1 and ℓ_2 , then w has a period of length $\gcd(\ell_1, \ell_2)$.

Lemma 2.14. [3] Let $\ell_1 \geq \ell_2$ and w be a word of length $n \geq \ell_1 + \ell_2$. If w has a period (or almost period) of length ℓ_1 and $w^{(k,m)}$ has a period of length ℓ_2 for some $m \in \mathcal{A}_d$, then there is $m' \in \mathcal{A}_d$ such that $w^{(k,m')}$ has a period of length $\gcd(\ell_1, \ell_2)$.

Lemma 2.15. Let $w \in \mathcal{A}_d^n$ such that w is ℓ_1 -periodic or almost ℓ_1 -periodic. Let $m \in [d - 1]$ and also $k \in [n]$ with $k \leq n - \ell_1$ or $k > \ell_1$. Then for any $\ell_2 < \frac{n}{2}$ with $\ell_1 \geq \ell_2$ and $n \geq \ell_1 + \ell_2$, it is not possible that $w^{(k,m)}$ is ℓ_2 -periodic.

Proof. We proceed by contradiction, and suppose that $w^{(k,m)}$ is ℓ_2 -periodic. We have two cases. First, if $k > \ell_1$, then by Lemma 2.14, there exists some $m' \in \mathcal{A}_d$ such that $w^{(k,m')}$ has period of length $\gcd(\ell_1, \ell_2)$. Then we have the following chain of equalities.

$$\begin{aligned} w_k &= w_{k-\ell_1} && \text{since } w \text{ has a period of length } \ell_1 \\ &= w_{k-\ell_2} && \text{since } w^{(k,m')} \text{ has a period of length } \gcd(\ell_1, \ell_2) \\ &= w_k^{(k,m)} && \text{since } w^{(k,m)} \text{ has a period of length } \ell_2 \end{aligned}$$

Hence this is a contradiction. For our second case, when $k \leq n - \ell_1$, we note the following.

$$w_k = w_{k+\ell_1} = w_{k+\ell_2} = w_k^{(k,m)}$$

This is also a contradiction. Therefore we must have that $w^{(k,m)}$ is not ℓ_2 -periodic. \square

Lemma 2.16. *Let $w \in \mathcal{A}_d^n$ such that w has period length ℓ_1 for some $\ell_1 < \frac{n}{2}$. For all $m \in [d-1]$ and for all $i, j, k \in [n]$ with $i \leq k \leq j$, and for all $\ell_2 \leq \ell_1$ with $j - i + 1 \geq \ell_1 + \ell_2$ and with either $k \geq i + \ell_1$ or $k \leq j - \ell_1$, we must have that $w^{(k,m)}(i, j)$ does not have period ℓ_2 .*

Proof. Define $w' = w(i, j)$ and $w'^{(k-i,m)} = w^{(k,m)}(i, j)$, and then apply Lemma 2.15 to compare the two strings. \square

Lemma 2.17. *Let $n = 2t$ and let $u \in \mathcal{A}_d^n$. If u has period length t and for some $\ell < t$ and $m \in \mathcal{A}_d$, we find that $u' = u^{(t,m)}(t+1-\ell : n-1)$ is ℓ -periodic, then we must have that ℓ divides t and $u' \oplus (u_n + m)$ has period ℓ .*

Proof. First, we note that $U_m = (u^{(t,m)})^{(n,m)}$ clearly has period length t , and so $U_m(t+1-\ell : n-1)$ has almost period length t . Additionally, since $U_m(t+1-\ell : n-1) = u'$, we know that $U_m(t+1-\ell : n-1)$ is ℓ -periodic. Hence by Lemma 2.13, $U_m(t+1-\ell : n-1)$ has period of length $p = \gcd(t, \ell)$. However since u' is given to be ℓ -periodic, the minimum period length is ℓ and so we must have that $p = \ell$ and thus ℓ divides t .

To show that $u' \oplus (u_n + m)$ has period ℓ , we note that u' has period ℓ , and that $u'_\ell = u_t^{(t,m)} = u_t + m$. Having period ℓ implies that $u'_k = u_t + m$ for all k that is divisible by ℓ . Since $u_n + m$ is the $(t+\ell)$ th letter in $u' \oplus (u_n + m)$, and this is divisible by ℓ , we need that $u_n + m = u_t + m$ in order for $u' \oplus (u_n + m)$ to have period ℓ . But this is given to be true since u has period t . \square

3 Identifying Codes

We begin by building up to the definition of an identifying code. This requires careful definitions of directed distance and t -balls.

Definition 3.1. The **directed distance** from vertex u to vertex v in a graph G is given by $\vec{d}(u, v)$, and is defined as the length of the shortest directed path from u to v in G .

Definition 3.2. Let $v \in V(G)$. The **open in-neighborhood** of v is given by $N^-(v) = \{u \in V(G) \mid (u, v) \in E(G)\}$, and the **closed in-neighborhood** is given by $N^-[v] = N^-(v) \cup \{v\}$.

Note that out-neighborhoods are defined analogously, but will not be needed in this paper.

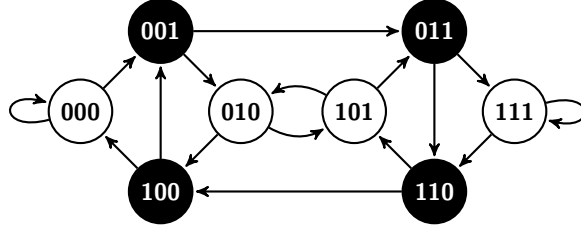


Figure 2: A 1-identifying code in the graph $\vec{\mathcal{B}}(2, 3)$ (black vertices). A 2-identifying code in this graph requires all vertices but one of $\{000, 111\}$, and there are no t -identifying codes for $t \geq 3$.

Definition 3.3. The **in-ball of radius t** centered at vertex v is the set: $B_t^-(v) = \{u \in V(G) \mid \vec{d}(u, v) \leq t\}$. and the **out-ball of radius t** centered at vertex v is the set: $B_t^+(v) = \{u \in V(G) \mid \vec{d}(v, u) \leq t\}$.

The two following lemmas are useful in working with distances in $\vec{\mathcal{B}}(d, n)$ and their proofs are self-evident.

Lemma 3.4. In $\vec{\mathcal{B}}(d, n)$ there is a directed path of length $t \leq n$ from x to y if and only if $x(t+1 : n) = y(1 : n-t)$. That is, if and only if the rightmost $n-t$ letters of x are the same as the leftmost $n-t$ letters of y .

Lemma 3.5. In $\vec{\mathcal{B}}(d, n)$ if vertices $x \neq y$ have the same prefix, then for all $u \neq \{x, y\}$, $\vec{d}(u, x) = \vec{d}(u, y)$. In particular, $B_t^-(x) \setminus \{x\} = B_t^-(y) \setminus \{y\}$ for all $t \leq n$.

Definition 3.6. Given a subset $S \subset V(G)$, the S **t -identifying set** for vertex v is given by $\text{ID}_S(v) = B_t^-(v) \cap S$.

Definition 3.7. A **t -identifying code** is set $S \subseteq V(G)$ such that each vertex has a unique, non-empty t -identifying set. That is, for every $u \in V(G)$, $\text{ID}_S(u) \neq \emptyset$, and for all pairs $u, v \in V(G)$ we have $\text{ID}_S(u) \neq \text{ID}_S(v)$. The variable t is referred to as the **radius** of the identifying code. See Figure 2 for an identifying code in the graph $\vec{\mathcal{B}}(2, 3)$.

In the above definitions, if t is omitted from the notation (i.e. identifying code instead of t -identifying code), then it is assumed that $t = 1$.

Note that not every graph has a t -identifying code for each t . In particular if the graph has two vertices with equal in-balls of radius t , then the graph

has no t -identifying code. The topic of such ‘ t -twins’ and the resulting non-existence of t -identifying codes is covered in Theorem 3.13.

Theorem 3.8. *If $\vec{\mathcal{B}}(d, n)$ is a t -identifiable graph, then the size of any t -identifying code is at least $d^{n-1}(d-1)$.*

Proof. Choose $t \leq n$ and $a \neq b$ in \mathcal{A} . Suppose that for some $w \in \mathcal{A}^{n-1}$, neither $x = w \oplus a$ nor $y = w \oplus b$ is in a set S . Since x and y share a prefix, by Lemma 3.5, $B_t^-(x) \setminus \{x\} = B_t^-(y) \setminus \{y\}$. Since neither x nor y is in S , $\text{ID}_S(x) = \text{ID}_S(y)$. Thus S is not a t -identifying code. Thus for each $w \in \mathcal{A}^{n-1}$, a t -identifying code must contain, at least, all but one of $w \oplus a$ for $a \in \mathcal{A}$. Thus a t -identifying code for $\vec{\mathcal{B}}(d, n)$ must have size at least $d^{n-1}(d-1)$. \square

Note that the result above is independent of the radius t . An interesting consequence of this is the fact that increasing the radius of our identifying code does not produce any decrease in the size of a minimum identifying code. For example, consider the potential application of identifying codes in sensor networks. One might think that by increasing the sensing power (which corresponds to the radius of the identifying code) we would be able to place fewer sensors and thus incur a savings overall. However, Theorem 3.8 implies that providing more powerful (and thus more expensive) sensors does not allow us to place fewer sensors. Thus we should use sensors that have sensing distance equivalent to radius one. In fact, in the case of 2-identifying codes in $\vec{\mathcal{B}}(2, 3)$, we actually require an extra vertex for a minimum size of seven!

The remainder of this section is organized as follows. We first provide a construction of an optimal t -identifying code for $\vec{\mathcal{B}}(d, n)$ with $t \geq 2$, and $n \geq 2t$ in Theorems 3.9 and 3.10. Following the proof of this result, we highlight some variations that provide identifying codes for several other instances. Finally, we highlight an alternative construction for 1-identifying codes for all $\vec{\mathcal{B}}(d, n)$ when $d \neq 2$ and n is odd.

Theorem 3.9. *Suppose that $n \geq 5, d \geq 2, t \geq 2$, and $n \geq 2t$. Then the following set S is an optimal t -identifying code of size $d^{n-1}(d-1)$ in $\vec{\mathcal{B}}(d, n)$.*

$$\begin{aligned}
S = & \{x \in \mathcal{A}_d^n \mid \text{for some } m \text{ and } \ell \leq t, x^{(t,m)}(t+1-\ell : n-1) \text{ is } \ell\text{-periodic,} \\
& \text{but } x^{(t,m)}(t+1-\ell : n-1) \oplus (x_n + m) \text{ is not.}\} \\
& \cup \\
& \{x \in \mathcal{A}_d^n \mid x_t \neq x_n \text{ and } x^{(t,m)}(t+1-\ell : n-1) \\
& \text{is not } \ell\text{-periodic for any } m \text{ and } \ell \leq t\}
\end{aligned}$$

Proof. First, let us note that for all $x \in \mathcal{A}_d^n$, if $x^{(t,m)}(t+1-\ell : n)$ has period ℓ for any m and $\ell \leq t$, then $x \notin S$.

Let x be given, and define x_i to be the i th coordinate of x . We also define the following string x_{ij} .

$$x_{ij}(k) = \begin{cases} i, & \text{if } k = t; \\ j, & \text{if } k = n; \\ x(k), & \text{otherwise.} \end{cases}$$

In other words, x_{ij} is the string $x_1 \dots x_{t-1} i x_{t+1} \dots x_{n-1} j$.

Next, we note that $x_{ij}(t+1-\ell : n-1) = x^{(t,i-x_t)}(t+1-\ell : n-1)$, which implies by Lemma 2.16 that $x_{ij}(t+1-\ell : n-1)$ is ℓ -periodic for at most one $\ell \leq t = \frac{n}{2}$ and for at most one $i \in \mathcal{A}_d$.

Next, suppose that $x_{ab}(t+1-\ell : n)$ is ℓ -periodic and consider x_{ij} . We note that $x_{ij}^{(t,a-i)}(t+1-\ell : n-1) = x_{ab}(t+1-\ell : n-1)$, which is ℓ -periodic. Hence x_{ij} is a member of S if and only if $x_{ij}^{(t,a-i)}(t+1-\ell : n-1) \oplus (j+a-i)$ is not ℓ -periodic. In other words, if and only if $i-j \not\equiv a-b \pmod{d}$. On the other hand, if $x_{ij}(t+1-\ell : n-1)$ is not ℓ -periodic for any ℓ and $i \in \mathcal{A}_d$, then $x_{ij} \in S$ if and only if $i \neq j$. Hence, there is exactly one $j \in \mathcal{A}_d$ such that $x_{aj} \notin S$, and similarly there is exactly one $i \in \mathcal{A}_d$ such that $x_{ib} \notin S$. These pairings tell us that d^{n-1} strings are not in S , leaving the cardinality of S at $d^n - d^{n-1}$, or $d^{n-1}(d-1)$.

Now that we have established the cardinality of S , we must show that no two nodes have the same identifying sets. Let $x, y \in \vec{\mathcal{B}}(d, n)$, and consider their identifying sets, called $I(x)$ and $I(y)$, respectively. Let k be the smallest index such that $x_k \neq y_k$.

$k = 1$: Without loss of generality, we may assume that $x_1 = 0$ and $y_1 = 1$.

Observe that we have the following strings contained in the identifying sets.

$$\begin{aligned} x' &= 0^{t-1} \oplus 0 \oplus x(1 : n-t) \in B_t^-(x), \\ x'' &= 0^{t-1} \oplus 1 \oplus x(1 : n-t) \in B_t^-(x), \\ y' &= 1^{t-1} \oplus 1 \oplus y(1 : n-t) \in B_t^-(y), \text{ and} \\ y'' &= 1^{t-1} \oplus 0 \oplus y(1 : n-t) \in B_t^-(y). \end{aligned}$$

Note that $x' \notin B_t^-(y)$ and $y' \notin B_t^-(x)$, since $B_t^-(x)$ does not contain any vertices beginning with 1^{t+1} and $B_t^-(y)$ does not contain any vertices beginning with 0^{t+1} . Next, we notice that at least one $\{x', x''\}$ is in S .

To see this, we note that $x'' = x''^{(t,1)}$. By the same point, we must have that at least one of y', y'' is a member of S .

Next, we note that if at least one of x', y' is a member of S , we can use that string to separate y . Otherwise, if either $x'' \notin B_t^-(y) \cap S$ or $y'' \notin B_t^-(x) \cap S$, we can separate x and y with the given string. As a last resort, we consider the case in which we have $x'' \in B_t^-(y) \cap S$ and $y'' \in B_t^-(x) \cap S$. Then we must have:

$$\begin{aligned} x'' &= 0^{t-1} \oplus y(1 : n - t + 1), \\ y'' &= 1^{t-1} \oplus x(1 : n - t + 1), \end{aligned}$$

as x''_t is the only 1 in $x''(1 : t + 1)$ and y''_t is the only 0 in $y''(1 : t + 1)$. From this, we get the following string equalities:

$$\begin{aligned} x(1 : n - t) &= y(2 : n - t + 1), \\ y(1 : n - t) &= x(2 : n - t + 1). \end{aligned}$$

From these string equalities, we see that $x_i = y_{i+1} = x_{i+2}$ and $y_i = x_{i+1} = y_{i+2}$ for all $i = 1, 2, \dots, n - t - 1$, and so $x(1 : n - t)$ and $y(1 : n - t)$ are both 2-periodic (since $x_1 \neq y_1$ we cannot have 1-periodic). Hence, $x''(t - 1 : n) = 0 \oplus 1 \oplus x(1 : n - t)$ is also 2-periodic (recall that $x_1 = 0$ and $y_1 = 1$), so $x'' \notin S$. Hence we must have $x' \in S$, and so we may use x' to separate x and y .

$2 \leq k \leq n - t$: We know that there must exist some s such that $x_1 \dots x_{s-1} = y_1 \dots y_{s-1}$, and these substrings are constant. Without loss of generality we may assume that $x_1 \dots x_{s-1} = 0^{s-1} = y_1 \dots y_{s-1}$ and $x_s = 1$, and so $2 \leq s \leq k$. Define the following strings.

$$\begin{aligned} y' &= 1^t \oplus y(1 : n - t) \text{ and} \\ y'' &= 1^{t-1} \oplus 0 \oplus y(1 : n - t) \end{aligned}$$

As we saw in the previous case, we have that $\{y', y''\} \subseteq B_t^-(y)$ and $\{y', y''\} \cap S \neq \emptyset$. Now consider an arbitrary vertex $v \in B_t^-(x)$. Since $x_{s-1}x_s = 01$ and $2 \leq s \leq k \leq n - t$, we know that $v(i - 1 : i) = 01$ for some $i \in [s, s + t]$.

Additionally, we consider y' and $i \in [2, s + t - 1]$. For $i \leq t$, we know that $y'(i - 1 : i) = 11$, and for $i = t + 1$, we have that $y'(i - 1 : i) = 10$, and

finally for $t+2 \leq i \leq s+t-1$ we have $y'(i-1:i) = 00$. Similarly, for $i \in [2, s+t-1]$, we must have $y''(i-1:i) \in \{00, 10, 11\}$. This implies that $y'(1:s+t-1)$ and $y''(1:s+t-1)$ do not contain the substring 01, and so if y' or y'' is a member of $B_t^-(x)$, we must have either $d(y', x) = t$ or $d(y'', x) = t$, respectively. Hence we must have $y'_{t+i} = x_i$ or $y''_{t+i} = x_i$ for $i \in [1, n-t]$, and therefore that $x_k = y'_{t+k} = y''_{t+k} = y_k$, which is a contradiction. Thus neither y' nor y'' can be a member of $B_t^-(x)$, and hence both strings separate x and y .

$k > n-t$: Since we must have $x_1 = y_1$, we may assume without loss of generality that $x_1 = y_1 \neq 0$. Define the following strings.

$$\begin{aligned} u &= 0^{n-k} \oplus x(1:k) \text{ and} \\ v &= 0^{n-k} \oplus y(1:k) \end{aligned}$$

Clearly we have $u \in B_t^-(x)$ and $v \in B_t^-(y)$. Additionally, we have $u(1:n-1) = v(1:n-1)$ and $u_n = x_k \neq y_k = v_n$. Note that this implies that for $a = x_{t-n+k}$ we have $u = u_{a,x_k}$ and $v = u_{a,y_k}$ with $x_k \neq y_k$. By our argument at the very beginning of the proof, at most one of these can lie outside of S , and so we must have $\{u, v\} \cap S \neq \emptyset$.

We now have two cases. First, if both $u \notin B_t^-(y)$ and $v \notin B_t^-(x)$, then any string from $\{u, v\} \cap S$ separates x and y , and we are done. Otherwise, assume without loss of generality that $u \in B_t^-(y)$. Then we must have $u = w \oplus y(1:p)$ for at least one $p \in [n-t, n]$. Take p to be the largest such p possible. Since $y_1 \neq 0 = u_i$ for all $i \in [1, n-k]$, we must have $p \leq k$. Additionally, if $p = k$ then we have $y(1:k) = y(1:p) = x(1:k)$, which is a contradiction since $y_k \neq x_k$. This implies that we must have $p < k$. Hence we must have the following string of equalities:

$$\begin{aligned} 0^{n-k} \oplus x(1:k) &= u \\ &= w \oplus y(1:p) \\ &= w \oplus x(1:p). \end{aligned}$$

Thus we have $x(k-p+1:k) = x(1:p)$, or $x_i = x_{k-p+i}$ for $i =$

$1, 2, \dots, p$. Additionally we note the following equalities hold:

$$\begin{aligned}
2(k-p) &= 2k - 2p \\
&\leq 2k - 2(n-t) \\
&\leq k + 2t - n \\
&\leq k.
\end{aligned}$$

Hence since $x_i = x_{k-p+i}$ for $i \in [1, p]$ and $2(k-p) \leq k$, we know that $x(1 : k)$ has period $(k-p)$. In fact, since we chose p to be maximum, $x(1 : k)$ is $(k-p)$ -periodic.

Next, we show that $u(t+1 - (k-p) : n)$ is also $(k-p)$ -periodic. First, we note the following inequalities:

$$\begin{aligned}
n - (t+1 - (k-p)) + 1 &= n - t + k - p \\
&\geq 2(k-p).
\end{aligned}$$

The last line comes from the facts: $n-t \geq t \geq n-p \geq k-p$. Hence our string length is at least $2(k-p)$, and from our previous paragraphs, so long as $u(t+1 - (k-p) : n)$ is contained in $u(n-k+1 : n) = x(1 : k)$, we know that it must have period $(k-p)$. For this we note that

$$t+1 - (k-p) \geq (n-p) + 1 - (k-p) = n-k+1,$$

and so $u(t+1 - (k-p) : n)$ indeed has period $(k-p)$, and thus $u \notin S$, except if $(n-1) - (t+1 - (k-p)) + 1 < 2(k-p)$. In this case we must have $k-p = t$, $k = n = 2t$, and $p = t$.

If $u^{(t,m)}(t+1-\ell : n-1)$ is ℓ -periodic for some $\ell < k-p = t$ and m , then by Lemma 2.17 we must have that $u^{(t,m)}(t+1-\ell : n-1) \oplus (u_n + m)$ is also ℓ -periodic, and hence $u \notin S$. On the other hand, if $u^{(t,m)}(t+1-\ell : n-1)$ is not ℓ -periodic for any $\ell < t$ and m , then we note that $u_n = u_{2t} = u_t$, so again $u \notin S$. Therefore in all cases we have $u \notin S$, so we must have $v \in S$.

All that remains is to show that $v \notin B_t^-(x)$. We note that if $v \in B_t^-(x)$, then $y(1 : k)$ is ℓ -periodic for some $\ell \leq k+t-n$ (using the same argument as we used to show that $x(1 : k)$ was $(k-p)$ -periodic). Since $x(1 : k) = (y(1 : k))^{(k,m)}$ for some m , by Lemma 2.15 it is not possible that both $x(1 : k)$ and $y(1 : k)$ are periodic. Hence we must have $v \notin B_t^-(x)$, and so we may use v to separate x and y .

□

When we combine the previous theorem with the following result, we have a complete set of constructions for optimal t -identifying codes in $\vec{\mathcal{B}}(d, n)$ with $t \geq 2$ and $n \geq 2t$.

Theorem 3.10. *Let $n > 3$ and $S = \mathcal{A}_d^n \setminus \{x_1 a x_3 x_4 \dots x_{n-1} a \mid a \in \mathcal{A}_d\}$. If n is even, then S is a 2-identifying code for $\vec{\mathcal{B}}(d, n)$. If n is odd, then $S' = (S \cup \{(ab)^{\frac{n-1}{2}} b \mid a \neq b \in \mathcal{A}_2\}) \setminus \{(ab)^{\frac{n-1}{2}} a \mid a \neq b \in \mathcal{A}_2\}$ is a 2-identifying code for $\vec{\mathcal{B}}(d, n)$. In both these cases, the 2-identifying code is of optimal size $d^{n-1}(d-1)$.*

Proof. Consider an arbitrary string $x = x_1 x_2 x_3 \dots x_n \in \mathcal{A}_d^n$, and define the set $T = \text{ID}_S(x)$. We'll consider the contents of T in four cases based on the equality of x_1, x_{n-1} and of x_2, x_n . First let $C = \{ax^{--} \mid a \in \mathcal{A} \setminus \{x_{n-2}\}\}$.

Case 1. If $x_2 = x_n$ and $x_1 = x_{n-1}$, then $T = \mathcal{A} \oplus C$. Thus $|T| = d^2 - d$.

Case 2. If $x_2 \neq x_n$ and $x_1 = x_{n-1}$, then $T = (\mathcal{A} \oplus C) \cup \{x\}$. If $x \in \mathcal{A} \oplus C$ then $x^+ = ax^{--}$ for some $a \in \mathcal{A} \setminus \{x_{n-2}\}$. In this case, we have $x_2 x_3 \dots x_n = ax_1 x_2 \dots x_{n-2}$. This implies that we have $x_1 = x_3 = x_5 = \dots$, and also that $x_2 = x_4 = x_6 = \dots$. Since this case requires that $x_1 = x_{n-1}$, we must have that either n is even or that $x_1 = x_2 = x_3 = \dots = x_n$. In either case, this contradicts our assumption that $x_2 \neq x_n$. Thus $x \notin \mathcal{A} \oplus C$, and we conclude that $|T| = d^2 - d + 1$.

Case 3. If $x_2 = x_n$ and $x_1 \neq x_{n-1}$, then $T = \mathcal{A} \oplus \{C \cup \{x^-\}\}$. If $x^- = ax^{--}$ for some $a \in \mathcal{A} \setminus \{x_{n-2}\}$, then $ax_1 x_2 \dots x_{n-2} = x_1 x_2 \dots x_{n-1}$. This implies that we have $x_1 = x_2 = x_3 = \dots = x_{n-2} = x_{n-1}$. This contradicts our assumption that $x_1 \neq x_{n-1}$. Thus $x^- \neq ax^{--}$ for any $a \in \mathcal{A} \setminus \{x_{n-2}\}$, and we conclude that $|T| = d^2$.

Case 4. If $x_2 \neq x_n$ and $x_1 \neq x_{n-1}$, then $T = (\mathcal{A} \oplus \{C \cup \{x^-\}\}) \cup \{x\}$. As in Case 3, since $x_1 \neq x_{n-1}$, $\mathcal{A} \oplus \{C \cup \{x^-\}\}$ contains d^2 distinct elements. Let us consider whether $x \in \mathcal{A} \oplus \{C \cup \{x^-\}\}$. If not, then $|T| = d^2 + 1$. There are two cases to consider.

a. If $x \in \mathcal{A} \oplus C$, then $x^+ = x^-$. In this case, we must have that $x_2 x_3 x_4 \dots x_n = x_1 x_2 \dots x_{n-1}$, which implies that we have the following chain of equalities: $x_1 = x_2 = x_3 = \dots = x_{n-1} = x_n$. This contradicts the assumptions that $x_2 \neq x_n$ and $x_1 \neq x_{n-1}$. Thus, this case does not occur.

b. If $x \in \mathcal{A} \oplus \{x^-\}$, then $x^+ = ax^{--}$ for some $a \in \mathcal{A}$. Then $x_2x_3 \cdots x_n = ax_1x_2 \cdots x_{n-2}$. This implies that $x_1 = x_3 = x_5 = \cdots$, and also that $x_2 = x_4 = x_6 = \cdots$. If n is even, this contradicts our assumptions that $x_2 \neq x_n$ and $x_1 \neq x_{n-1}$. Thus for even n , this case does not occur. For n odd, this case only occurs if $x \in \{(ab)^{\frac{n-1}{2}}a\}$.

Thus, if n is even, or n is odd and $x \notin \{(ab)^{\frac{n-1}{2}}a\}$, we can see that $T = \text{ID}_S(x)$ completely determines the string x . In particular, given T we can decide which case we are in based on $|T|$. We can then determine x_1, \dots, x_n based on the content of T . Thus in these cases S is an identifying code.

However, if n is odd, and $x \in \{(ab)^{\frac{n-1}{2}}a\}$ we must change S to get an identifying code. Note that $B_2^-((ab)^{\frac{n-1}{2}}a) \cup \{(ab)^{\frac{n-1}{2}}b\} = B_2^-((ab)^{\frac{n-1}{2}}b)$. Since our set S contains vertices of the form $(ab)^{\frac{n-1}{2}}a$ but not $(ab)^{\frac{n-1}{2}}b$, these two types of vertices must have identical identifying sets with respect to S . Thus by adding the vertices in $\{(ab)^{\frac{n-1}{2}}b\}$, we are able to create distinct identifying sets with respect to $S \cup \{(ab)^{\frac{n-1}{2}}b\}$. However, we note that we now have the vertices of $\{(ab)^{\frac{n-1}{2}}b\}$ and $\{(ab)^{\frac{n-1}{2}}a\}$ in our identifying code, but that $B_2^+((ab)^{\frac{n-1}{2}}a) \cup \{b(ba)^{\frac{n-1}{2}}\} = B_2^+(b(ba)^{\frac{n-1}{2}})$. This implies that the inclusion of both $(ab)^{\frac{n-1}{2}}b$ and $(ab)^{\frac{n-1}{2}}a$ in our identifying code is only necessary if they are required to identify vertex $(ab)^{\frac{n-1}{2}}a$ from vertex $b(ba)^{\frac{n-1}{2}}$. So, as long as we can identify $(ab)^{\frac{n-1}{2}}a$ differently from $b(ba)^{\frac{n-1}{2}}$ without using $b(ba)^{\frac{n-1}{2}}$, we need only include $(ab)^{\frac{n-1}{2}}b$ and not $(ab)^{\frac{n-1}{2}}a$ in our identifying code. Since these two vertices have disjoint in-balls of radius 2 for $n > 3$, they must have distinct 2-identifying sets. Thus S' is a 2-identifying code in this case. \square

Finally, we provide additional constructions of identifying codes. Theorems 3.11 and 3.12 have proofs very similar to that of Theorem 3.9, so we omit them here.

Theorem 3.11. *Assume that $d \geq 2$, and $n \geq 3$. Then the following subset S is an optimal 1-identifying code of size $d^{n-1}(d-1)$ in $\vec{\mathcal{B}}(d, n)$.*

$$\begin{aligned}
S = & \{x \in \mathcal{A}_d^n \mid \text{for some } m \text{ and } \ell \in \{1, 2\}, x^{(1,m)}(1 : n-1) \text{ is } \ell\text{-periodic or} \\
& \text{almost } \ell\text{-periodic, but } x^{(1,m)}(1 : n-1) \oplus (x_n + m) \text{ is not.}\} \\
& \cup \\
& \{x \in \mathcal{A}_d^n \mid x_1 \neq x_n \text{ and } x^{(1,m)}(1 : n-1) \\
& \text{is not } \ell\text{-periodic for any } m \text{ and } \ell \in \{1, 2\}\}
\end{aligned}$$

Theorem 3.12. *Assume that $d \geq 2$. Then the following subset S is a t -identifying code of size $d^{n-1}(d-1)+d^t$ in the directed de Bruijn graph $\vec{\mathcal{B}}(d, n)$, if $n = 2t - 1 \geq 5$.*

$$\begin{aligned}
S = & \{x \in \mathcal{A}_d^n \mid \text{for some } m \text{ and } l < t - 1: x^{(t,m)}(t+1-l : n-1) \text{ is } l\text{-periodic,} \\
& \text{but } x^{(t,m)}(t+1-l : n-1) \oplus (x_n + m) \text{ is not.}\} \\
& \cup \\
& \{x \in \mathcal{A}_d^n \mid x_t \neq x_n \text{ and } x^{(t,m)}(t+1-l : n-1) \text{ is not } l\text{-periodic} \\
& \text{for any } m \text{ and } l < t - 1.\} \\
& \cup \\
& \{x \in \mathcal{A}_d^n \mid x^{(t,m)}(1 : n-1) \text{ is almost } t\text{-periodic, for some } m.\}
\end{aligned}$$

We note that the construction in Theorem 3.12 is not optimal. To find an optimal t -identifying code when $n = 2t - 1$ is an open problem to be considered in the future. For the cases when $n < 2t - 1$, we have the following theorem.

Theorem 3.13. *There is no t -identifying code in the directed de Bruijn graph $\vec{\mathcal{B}}(d, n)$ when $n \leq 2t - 2$.*

Proof. Let $u = 0^{n-t} \oplus 1 \oplus 0^{t-2} \oplus 1$ and $v = 0^{n-t} \oplus 1 \oplus 0^{t-2} \oplus 0$. Since $B_t^-(u)$ and $B_t^-(v)$ contain all vertices that end with 0^{n-t} or $0^{n-t} \oplus 1 \oplus 0^k$ where $k = 0, 1, \dots, n-t-1$, u and v are t -twins. Thus $\vec{\mathcal{B}}(d, n)$ has no t -identifying code. \square

As an additional treat for the reader, we provide a simple construction for 1-identifying codes in $\vec{\mathcal{B}}(d, n)$ whenever we have either $d > 2$ or n odd.

Theorem 3.14. *If n is odd, or n is even and $d > 2$, then*

$$S = \mathcal{A}_d^n \setminus \{a \oplus \mathcal{A}_d^{n-2} \oplus a \mid a \in \mathcal{A}_d\}$$

is an identifying code for $\vec{\mathcal{B}}(d, n)$. Further this identifying code has optimal size $(d-1)d^{n-1}$.

Proof. Define S as in the statement of the theorem. First, we will see that the identifying set for every vertex has size either d or $d-1$. Let $x = x_1x_2 \dots x_n$, then

$$N^-(x) \cap S = \{\mathcal{A} \oplus x_1x_2 \dots x_{n-1}\} \setminus \{x_{n-1}x_1x_2 \dots x_{n-1}\}.$$

If $x_1 = x_n$, then $\text{ID}_S(x) = N^-(x) \cap S$ has size $d - 1$. Whereas, if $x_1 \neq x_n$, then $\text{ID}_S(x) = \{x\} \cup N^-(x) \cap S$ has size d .

From this it is clear that every vertex has a non-empty identifying set. However we must also show that every identifying set is unique. Suppose there are two distinct vertices $x, y \in V(\vec{\mathcal{B}}(d, n))$ such that $\text{ID}_S(x) = \text{ID}_S(y)$. Call their identical identifying set T . We look at the two cases, $|T| = d$ and $|T| = d - 1$, separately below.

Suppose that $|T| = d$. Then $\{x, y\} \subseteq T$ by our assumption on T and our earlier reasoning. Since $x \neq y$, this means that $\vec{\mathcal{B}}(d, n)$ contains both directed arcs $x \rightarrow y$ and $y \rightarrow x$. This allows us to conclude that $\{x, y\} = \{(ab)^k, (ba)^k\}$ for some distinct $a, b \in \mathcal{A}$ with $k = n/2$. In particular we must have n even. Below are the precise identifying sets for x and y .

$$\begin{aligned}\text{ID}_S((ab)^k) &= \{(ab)^k, (ba)^k\} \cup \{c(ab)^{k-1}a \mid c \in \mathcal{A} \setminus \{a, b\}\} \\ \text{ID}_S((ba)^k) &= \{(ab)^k, (ba)^k\} \cup \{c(ba)^{k-1}b \mid c \in \mathcal{A} \setminus \{a, b\}\}\end{aligned}$$

If $d > 2$ these two identifying sets are in fact different, which is a contradiction.

Suppose that $|T| = d - 1$. Then neither x nor y is in T , which means neither is in S . However since their identifying sets are identical, this means that they have identical first neighborhoods. By definition of first neighborhoods, this means that x and y have the same prefix but different final letters. By then definition of S , one of x, y (if not both) is a member of S , which is a contradiction. \square

4 Dominating, Resolving, and Determining Sets

In this section we examine other types of vertex sets which identify or classify vertices up to some graph property. The properties used to define these sets are adjacency, distance, and automorphisms.

4.1 Dominating Sets

Definition 4.1. A (directed) **t -dominating set** is a subset $S \subseteq V(G)$ such that for all $v \in V(G)$ we have $B_t^-(v) \cap S \neq \emptyset$. That is, S is a (directed)

t -dominating set if every vertex in G is within (directed) distance t of some vertex in S . We denote the size of a minimum t -dominating set in a graph G by $\gamma_t(G)$.

Note that by definition every identifying code is also a dominating set, but not conversely.

Theorem 4.2. [11] For $d \geq 2$, $n \geq 1$, $\gamma_1(\vec{\mathcal{B}}(d, n)) = \lceil \frac{d^n}{d+1} \rceil$.

In [11] a construction of a minimum dominating set for $\vec{\mathcal{B}}(d, n)$ is given. Key to this construction is the fact that every integer m corresponds to a string (base d) in \mathbb{Z}_d^n , that we call X_m . The construction utilizes a special integer m defined by:

$$m = \begin{cases} d^{n-2} + d^{n-4} + \dots + d^{n-2k} + \dots + d^2 + 1 \bmod d^n, & \text{if } n \text{ is even;} \\ d^{n-2} + d^{n-4} + \dots + d^{n-2k} + \dots + d^3 + d \bmod d^n, & \text{if } n \text{ is odd.} \end{cases}$$

Let $D = \{m, m+1, \dots, m + \lceil \frac{d^n}{d+1} \rceil - 1\}$. Now let S be the set of strings $\{X_i \mid i \in D\}$. Then S is a minimum size dominating set for $\vec{\mathcal{B}}(d, n)$.

Next we provide constructions for t -dominating sets. While others have considered some variations of t -dominating sets (such as perfect dominating sets in [12]), it does not appear that the general t -dominating sets have been considered in the directed de Bruijn graph.

Theorem 4.3. The set $S \cup \{0^n\}$ where

$$S = \{x \in \mathcal{A}_d^n \mid x_{k(t+1)} \neq 0 \text{ for some } k \in \mathbb{Z}_+ \text{ and } x_i = 0 \text{ for all } i < k(t+1)\}$$

is a t -dominating set of size

$$1 + d^{n-t-1}(d-1) \left(\frac{1 - d^{-(t+1)\lfloor \frac{n}{t+1} \rfloor}}{1 - d^{-(t+1)}} \right)$$

in $\vec{\mathcal{B}}(d, n)$.

Proof. Let x be a vertex in \mathcal{A}_d^n . Assume that there are k zeros at the beginning of x , but not $k+1$ zeros, i.e. $x = 0^n$ or $x = 0^k \oplus a \oplus x(k+2 : n)$ for

some $a \neq 0$. Let $l \in [0, t]$ be an integer so that $k + l \equiv t \pmod{t+1}$, i.e. $k + l = t + m(t+1) = (m+1)(t+1) - 1$ for some $m \geq 0$. Now

$$0^l \oplus x(1 : n-l) = 0^l \oplus 0^k \oplus a \oplus x(k+2 : n-l) = 0^{k+l} \oplus a \oplus x(k+2 : n-l)$$

belongs to S except if $k+l \geq n$. If $k+l \geq n$, then $0^{n-k} \oplus x(1 : k) = 0^n \in S \cup \{0^n\}$ dominates x . Therefore every vertex is dominated by $S \cup \{0^n\}$.

There are $d^{n-k(t+1)} \cdot (d-1)$ vertices which begin with exactly $k(t+1) - 1$ zeros. Moreover, every vertex of $S \setminus \{0^n\}$ begins at most $n-1$ zeros. This needs that $k(t+1) < n$ or $1 \leq k \leq \lfloor \frac{n}{t+1} \rfloor$. Finally, 0^n is added to the dominating set $S \cup \{0^n\}$. Therefore the size of $S \cup \{0^n\}$ is

$$\begin{aligned} & 1 + \sum_{i=1}^{\lfloor \frac{n}{t+1} \rfloor} d^{n-i(t+1)} \cdot (d-1) \\ &= 1 + d^n(d-1) \sum_{i=1}^{\lfloor \frac{n}{t+1} \rfloor} (d^{-t-1})^i \\ &= 1 + d^n(d-1) \left(-1 + \sum_{i=0}^{\lfloor \frac{n}{t+1} \rfloor} (d^{-t-1})^i \right) \\ &= 1 + d^n(d-1) \left(-1 + \frac{1 - (d^{-t-1})^{\lfloor \frac{n}{t+1} \rfloor + 1}}{1 - d^{-(t+1)}} \right) \\ &= 1 + d^n(d-1) \left(\frac{d^{-(t+1)} - (d^{-(t+1)})^{\lfloor \frac{n}{t+1} \rfloor + 1}}{1 - d^{-(t+1)}} \right) \\ &= 1 + d^{n-t-1}(d-1) \left(\frac{1 - d^{-(t+1)\lfloor \frac{n}{t+1} \rfloor}}{1 - d^{-(t+1)}} \right). \end{aligned}$$

□

This result gives us the following lower bound on the size of a t -dominating set in $\vec{\mathcal{B}}(d, n)$.

Theorem 4.4. *Bounds on the size of a t -dominating set in $\vec{\mathcal{B}}(d, n)$ are given*

by:

$$\gamma_t(\vec{\mathcal{B}}(d, n)) \geq \begin{cases} 1 + d^{n-t-1}(d-1) \left(\frac{1-d^{-(t+1)} \lfloor \frac{n}{t+1} \rfloor}{1-d^{-(t+1)}} \right) & \text{if } n \equiv t \pmod{t+1} \\ d^{n-t-1}(d-1) \left(\frac{1-d^{-(t+1)} \lfloor \frac{n}{t+1} \rfloor}{1-d^{-(t+1)}} \right) & \text{otherwise.} \end{cases}$$

Proof. Suppose that the set S' is a t -dominating set in $\vec{\mathcal{B}}(d, n)$. Choose $a \in \mathcal{A}_d \setminus \{0\}$ and $i \in \mathbb{Z}$ so that $i \leq \frac{n}{t+1}$, and $w \in \mathcal{A}_d^{n-i(t+1)}$. Let $x = w \oplus a \oplus 0^{i(t+1)-1}$. We note that $B_t^-(x)$ contains the following elements.

$$B_t^-(x) = \{y \mid y = w' \oplus w \oplus a \oplus 0^{i(t+1)-1-k} \text{ for some } k \in [0, t], w' \in \mathcal{A}_d^k\}$$

Note that for all $v \neq x$ such that $v = v' \oplus b \oplus 0^{j(t+1)-1}$ with $b \in [d-1]$, $j \leq \frac{n}{t+1}$, and $v' \in \mathcal{A}_d^{n-j(t+1)}$, we must have that $B_t^-(x) \cap B_t^-(v) = \emptyset$. Hence each of these types of strings must be dominated by a different element of S' , and so we must have the following lower bound on $|S'|$. Define $A = \{v \mid v = v' \oplus b \oplus 0^{j(t+1)-1} \text{ with } b \in [d-1], j \leq \frac{n}{t+1}, v' \in \mathcal{A}_d^{n-j(t+1)}\}$.

$$\begin{aligned} |S'| &\geq |A| \\ &= \sum_{j=1}^{\lfloor \frac{n}{t+1} \rfloor} d^{n-j(t+1)} \cdot (d-1) \\ &= d^{n-t-1}(d-1) \left(\frac{1-d^{-(t+1)} \lfloor \frac{n}{t+1} \rfloor}{1-d^{-(t+1)}} \right) \end{aligned}$$

Finally, we consider the string 0^n and note that

$$B_t^-(0^n) = \{z \mid z = z' \oplus 0^{n-s} \text{ with } z' \in \mathcal{A}_d^s, s \leq t\}.$$

When we compare $B_t^-(0^n)$ with $B_t^-(x)$, we note that since $a \neq 0$ we must have that the closest element of $B_t^-(x)$ to 0^n is x itself. Next we note that the string closest to 0^n in the set A will occur when $j = \lfloor \frac{n}{t+1} \rfloor$. This will give us the string with the most 0's packed at the right end. Finally, if $n \equiv p \pmod{t+1}$, then this string looks like $v' \oplus b \oplus 0^{n-p-1}$ with $v' \in \mathcal{A}_d^p$ and $b \neq 0$. If $p = t$, then we are still unable to reach 0^n , and so we must have at least one additional string in S' to cover 0^n . \square

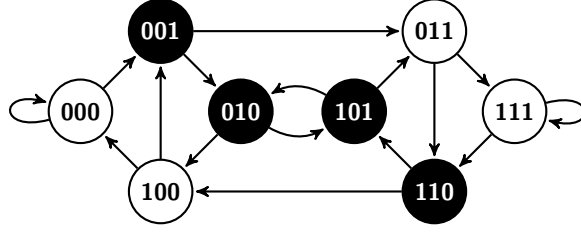


Figure 3: A directed resolving set in the graph $\vec{\mathcal{B}}(2, 3)$ (black vertices).

4.2 Resolving Sets

Definition 4.5. A **directed resolving set** is a set S such that for all $u, v \in V(G)$ there exist $s \in S$ so that $\vec{d}(s, u) \neq \vec{d}(s, v)$. The **directed metric dimension** is the minimum size of a directed resolving set. An example of a directed resolving set in $\vec{\mathcal{B}}(2, 3)$ is given in Figure 3.

Note that this definition is not quite the same as that given in [6] (which requires that there exist $s \in S$ so that $\vec{d}(u, s) \neq \vec{d}(v, s)$). Our definition corresponds better to the definitions of domination and of identifying codes for directed graphs that are used in this paper.

Theorem 4.6. *The directed metric dimension for $\vec{\mathcal{B}}(d, n)$ is $d^{n-1}(d-1)$.*

Proof. The following shows that for each $w \in \mathcal{A}^{n-1}$ a directed resolving set for $\vec{\mathcal{B}}(d, n)$ must contain (at least) all but one of the vertices with prefix w . Suppose that $w \in \mathcal{A}^{n-1}$, and $i \neq j \in \mathcal{A}$ so that neither of $w \oplus i, w \oplus j$ is in our set S . Note that if $x, y \in V(\vec{\mathcal{B}}(d, n))$, with $x \neq y$, then the distance from x to y is completely determined by x^- (and y^+). Since neither $w \oplus i$ nor $w \oplus j$ is in S , and both have the same prefix, $\vec{d}(w \oplus i, x) = \vec{d}(w \oplus j, x)$ for all $x \in S$. Thus S is not a directed resolving set. Thus for every $w \in \mathcal{A}^{n-1}$, S must contain (at least) all but one of the strings $w \oplus j$ for $j \in \mathcal{A}$. Thus $|S| \geq d^{n-1}(d-1)$. Since $\{w \oplus 0 \mid w \in \mathcal{A}^{n-1}\}$ can easily be shown to be a directed resolving set, we have the desired equality. \square

The combination of Theorem 3.8 and Theorem 4.2 yields:

Corollary 4.7. *The directed metric dimension for $\vec{\mathcal{B}}(d, n)$ is equal to the minimum size of a t -identifying code for $\vec{\mathcal{B}}(d, n)$ if $2t \leq n$.*

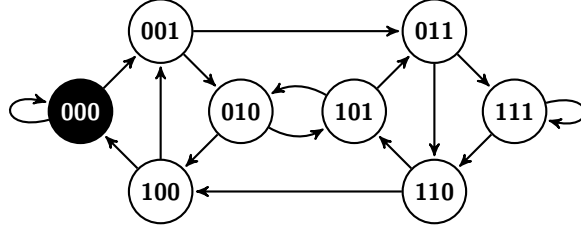


Figure 4: A minimum size determining set for $\vec{\mathcal{B}}(2, 3)$ (black vertex).

4.3 Determining Sets

In this section we will use a determining set to help us illustrate the automorphism group of $\vec{\mathcal{B}}(d, 2)$, study the relationship between $\text{Aut}(\vec{\mathcal{B}}(d, n - 1))$ and $\text{Aut}(\vec{\mathcal{B}}(d, n))$ and use the result to find the determining number for each $\vec{\mathcal{B}}(d, n)$. First let's recall some definitions.

Definition 4.8. An **automorphism** of a graph G is a permutation π of the vertex set such that for all pairs of vertices $u, v \in V(G)$, uv is an edge between u and v if and only if $\pi(u)\pi(v)$ is an edge between $\pi(u)$ and $\pi(v)$. An **automorphism** of a directed graph G is a permutation π of the vertex set such that for all pairs of vertices $u, v \in V(G)$, uv is an edge from u to v if and only if $\pi(u)\pi(v)$ is an edge from $\pi(u)$ to $\pi(v)$. One automorphism in the binary (directed or undirected) de Bruijn graph is a map that sends each string to its complement.

Definition 4.9. [4] A **determining set** for G is a set S of vertices of G with the property that the only automorphism that fixes S pointwise is the trivial automorphism. The **determining number** of G , denoted $\text{Det}(G)$ is the minimum size of a determining set for G . See Figure 4 for an example.

Note that an alternate definition for a determining set is a set S with the property that whenever $f, g \in \text{Aut}(G)$ so that $f(s) = g(s)$ for all $s \in S$, then $f(v) = g(v)$ for all $v \in V(G)$. That is, every automorphism is completely determined by its action on a determining set.

Notice that since for both directed resolving sets and for identifying codes, since each vertex in a graph is *uniquely* identified by its relationship to the subset by properties preserved by automorphisms, the subset is also a determining set. Thus every directed resolving set and every identifying code

is a determining set. However, though domination is preserved by automorphisms, vertices are not necessarily uniquely identifiable by their relationship to a dominating set. Thus a dominating set is not necessarily a determining set. However, the relationships above mean that the size of a minimum determining set must be at most the size of a minimum identifying code or the directed metric dimension. For de Bruijn graphs we have shown that the latter numbers are rather large. Does this mean that the determining number is also large. We will see in Corollary 4.13 that the answer for directed de Bruijn graphs is a resounding ‘No’.

Lemma 4.10. $S = \{00, 11, 22, 33, \dots, (d-1)(d-1)\}$ is a determining set for $\vec{\mathcal{B}}(d, 2)$.

Proof. Suppose that $\sigma \in \text{Aut}(\vec{\mathcal{B}}(d, 2))$ fixes S pointwise. That is, $\sigma(ii) = ii$ for all $i \in \mathcal{A}$. Choose $ij \neq rs \in V(\vec{\mathcal{B}}(d, 2))$. Then either $i \neq r$ or $j \neq s$ (or both). If $i \neq r$ then $\vec{d}(ii, rs) = 2$ which is distinct from $\vec{d}(ii, ij) = 1$. Since an automorphism of a directed graph must preserve directed distance, $\sigma(ij) \neq rs$ if $i \neq r$. If $j \neq s$, then $\vec{d}(rs, jj) = 2$ which is distinct from $\vec{d}(ij, jj) = 1$. Thus, again using that σ preserves directed distance, $\sigma(ij) \neq rs$ if $j \neq s$. Thus, $\sigma(ij) = ij$ for all $ij \in V(\vec{\mathcal{B}}(d, 2))$ and therefore σ is the identity map and S is a determining set. \square

Note that we are using directed distances both from and to elements of the set S . Thus S does not fit the definition of a directed resolving set for $\vec{\mathcal{B}}(d, 2)$ (by [6], this would require that each vertex $v \in V(G)$ be distinguished by its directed distance to the vertices of the resolving set). However directed distances both to and from a set can be used in determining automorphisms of a directed graph.

Lemma 4.11. $\text{Aut}(\vec{\mathcal{B}}(d, 2)) \cong \text{Sym}(\mathcal{A}_d)$.

Proof. Let $\sigma \in \text{Sym}(\mathcal{A}_d)$. Define φ_σ on $V(\vec{\mathcal{B}}(d, n))$ by applying σ to each vertex coordinate-wise. That is $\varphi_\sigma(ab) = \sigma(a)\sigma(b)$. It is easy to show that φ_σ preserves directed edges and thus is an automorphism. Further, distinct permutations in $\text{Sym}(\mathcal{A}_d)$ produce distinct automorphisms since they act differently on the vertices of the determining set S defined above. Thus we have an injection $\text{Sym}(\mathcal{A}_d) \hookrightarrow \text{Aut}(\vec{\mathcal{B}}(d, 2))$.

Since the vertices of S are precisely the vertices with loops, every automorphism of $\vec{\mathcal{B}}(d, 2)$, must preserve S setwise. This provides the necessary injection from $\text{Aut}(\vec{\mathcal{B}}(d, 2)) \hookrightarrow \text{Sym}(\mathcal{A})$. Thus, $\text{Aut}(\vec{\mathcal{B}}(d, 2)) \cong \text{Sym}(\mathcal{A}_d)$. \square

Note that we can consider the automorphisms of $\vec{\mathcal{B}}(d, 2)$ as permutations of the loops, but we can simultaneously consider them as permutations of the symbols in the alphabet \mathcal{A}_d . It can be useful to view the automorphisms in these two different ways.

Note that as shown in [1], $\vec{\mathcal{B}}(d, n)$ can be built inductively from $\vec{\mathcal{B}}(d, n-1)$ in the following way. The vertex $x_1 \dots x_n \in V(\vec{\mathcal{B}}(d, n))$ corresponds to the directed edge from $x_1 \dots x_{n-1}$ to $x_2 \dots x_n$ in $\vec{\mathcal{B}}(d, n-1)$. The directed edge $\vec{\mathcal{B}}(d, n)$ from $x_1 \dots x_n \rightarrow x_2 \dots x_n x_{n+1}$ corresponds to the directed 2-path $x_1 \dots x_{n-1} \rightarrow x_2 \dots x_n \rightarrow x_3 \dots x_{n+1}$ in $\vec{\mathcal{B}}(d, n-1)$. That is, $\vec{\mathcal{B}}(d, n)$ is the directed line graph of the directed graph $\vec{\mathcal{B}}(d, n-1)$. Thus, by [8] (Chapter 27, Section 1.1), $\text{Aut}(\vec{\mathcal{B}}(d, n-1)) = \text{Aut}(\vec{\mathcal{B}}(d, n)) \cong \text{Sym}(\mathcal{A})$. In the following paragraphs we see detail this correspondence.

Suppose that $\varphi \in \text{Aut}(\vec{\mathcal{B}}(d, n))$. Since φ preserves directed edges, we know that both $\varphi(x_1 \dots x_n) = a_1 \dots a_n$ and $\varphi(x_2 \dots x_{n+1}) = b_1 \dots b_n$ if and only if $a_2 = b_1, \dots, a_n = b_{n-1}$. Thus if $\varphi(x_1 \dots x_{n-1} x_n) = a_1 \dots a_{n-1} a_n$ then for every $b \in \mathcal{A}$, $\varphi(x_1 \dots x_{n-1} z) = a_1 \dots a_{n-1} c$ for some $c \in \mathcal{A}$. In particular, this allow us to define an automorphism $\varphi' \in \text{Aut}(\vec{\mathcal{B}}(d, n-1))$ corresponding to $\varphi \in \text{Aut}(\vec{\mathcal{B}}(d, n))$. Define φ' by $\varphi'(x_1 \dots x_{n-1}) = a_1 \dots a_{n-1}$ where $\varphi(x_1 \dots x_n) = a_1 \dots a_{n-1} a_n$. By the preceding discussion, φ' is well-defined. It is also clearly a bijection on vertices of $\vec{\mathcal{B}}(d, n-1)$. Consider $x_1 \dots x_{n-1}$ and $x_2 \dots x_{n-1} x_n$, the initial and terminal vertices of a directed edge in $\vec{\mathcal{B}}(d, n-1)$. Since φ preserves directed edges if $\varphi(x_1 \dots x_n) = a_1 \dots a_{n-1} a_n$ then for any $z \in \mathcal{A}$, $\varphi(x_2 \dots x_n z) = a_2 \dots a_n w$ for some $w \in \mathcal{A}$. By definition of φ' , $\varphi'(x_1, \dots, x_{n-1}) = a_1 \dots a_{n-1}$ and $\varphi'(x_2 \dots x_n) = a_2 \dots a_n$. Thus φ' preserves the directed edge. Thus we get $\text{Aut}(\vec{\mathcal{B}}(d, n)) \hookrightarrow \text{Aut}(\vec{\mathcal{B}}(d, n-1))$.

In the other direction, suppose we are given $\varphi' \in \text{Aut}(\vec{\mathcal{B}}(d, n-1))$. Since φ' preserves directed edges, and directed edges of $\vec{\mathcal{B}}(d, n-1)$ are precisely the vertices of $\vec{\mathcal{B}}(d, n)$, φ' defines a map on vertices of $\vec{\mathcal{B}}(d, n)$. That is, (with some abuse of notation)

$$\begin{aligned} \varphi(x_1 \dots x_n) &= \varphi(x_1 \dots x_{n-1} \rightarrow x_2 \dots x_n) \\ &= \varphi'(x_1 \dots x_{n-1} \rightarrow x_2 \dots x_n) \\ &= \varphi'(x_1 \dots x_{n-1}) \rightarrow \varphi'(x_2 \dots x_n). \end{aligned}$$

Thus, given $\varphi'(x_1 \dots x_{n-1}) = a_1 \dots a_{n-1}$ then $\varphi'(x_2 \dots x_n) = a_2 \dots a_n$ for some $a_n \in \mathcal{A}$ and we define $\varphi(x_1 \dots x_n) = a_1 \dots a_n$. Further, since φ' pre-

serves directed 2-paths, φ preserves directed edges. Thus we get

$$\text{Aut}(\vec{\mathcal{B}}(d, n-1)) \hookrightarrow \text{Aut}(\vec{\mathcal{B}}(d, n)).$$

Since the automorphisms of $\vec{\mathcal{B}}(d, 2)$ are permutations of the loops, and of the symbols of \mathcal{A} , by induction, so are the automorphisms of $\vec{\mathcal{B}}(d, n)$ for all n . Thus we have proved the following.

Theorem 4.12. $\text{Aut}(\vec{\mathcal{B}}(d, n)) \cong \text{Sym}(\mathcal{A}_d)$ for all $n \geq 2$.

Corollary 4.13. $\text{Det}(\vec{\mathcal{B}}(d, n)) = \lceil \frac{d-1}{n} \rceil$.

Proof. Let S be a minimum set of vertices in which each letter of \mathcal{A}_{d-1} occurs at least once. It is easy to see that $|S| = \lceil \frac{d-1}{n} \rceil$. Any permutation of \mathcal{A}_d that acts nontrivially on any letter of \mathcal{A}_d must act non-trivially on any string containing that letter. Thus if $\sigma \in \text{PtStab}(S)$, then σ must fix every letter contained in any string in S . Thus σ fixes $0, 1, \dots, d-1$ and therefore also d . We can conclude that σ is the identity in both $\text{Sym}(\mathcal{A}_d)$ and in $\text{Aut}(\vec{\mathcal{B}}(d, n))$ and therefore S is a determining set. Thus $\text{Det}(\vec{\mathcal{B}}(d, n)) \leq \lceil \frac{d-1}{n} \rceil$.

Further if $|S| < \lceil \frac{d-1}{n} \rceil$ then fewer than $d-1$ letters of \mathcal{A}_d are used in strings in S . If $a, b \in \mathcal{A}_d$ are not represented in S , then the transposition $(a\ b)$ in $\text{Sym}(\mathcal{A}_d)$ is a non-trivial automorphism of $\vec{\mathcal{B}}(d, n)$ that fixes S pointwise. Thus S is not a determining set. \square

Thus for directed de Bruijn graphs, the determining number and the directed metric dimension can be vastly different in size.

5 Future Work

There are several directions that future work in this research area could take. The first is to continue the research on identifying codes in directed de Bruijn graphs. One key result missing from this paper is the determination of the size of a minimum t -identifying code in the graph $\vec{\mathcal{B}}(d, n)$ when $t = 2n - 1$. It would be ideal to determine both a formula for which graphs $\vec{\mathcal{B}}(d, n)$ are t -identifiable for a given t , as well as constructions when it is known that such an identifying code exists.

An alternative direction for future research is to consider these same vertex subsets (identifying codes, dominating sets, resolving sets, and determining sets) on the undirected de Bruijn graph. Little work has been

done and even foundational results like the size of a minimum dominating set are currently missing. Basic Matlab programs have shown that many more undirected de Bruijn graphs are t -identifiable than directed, and also that the minimum size of an identifying code is much smaller. For example, through brute force testing we have determined that the minimum size of a 1-identifying code in $\mathcal{B}(2, 5)$ is 12, whereas in the directed graph $\vec{\mathcal{B}}(2, 5)$ we have shown that the minimum size is 16.

Finally, variations on the concept of identifying code would be useful for real-world applications. For example, one type of variation known as a k -robust identifying code allows for up to k sensor (identifying code vertex) failures without disruption of the identifying code properties.

References

- [1] Toru Araki. On the k -tuple domination of de Bruijn and Kautz digraphs. *Inform. Process. Lett.*, 104(3):86–90, 2007. [4.3](#)
- [2] Joel Baker. De Bruijn graphs and their applications to fault tolerant networks. Master’s thesis, Oregon State University, 2011. [1](#)
- [3] J. Berstel and L. Boasson. Partial words and a theorem of Fine and Wilf. *Theoret. Comput. Sci.*, 218:135–141, 1999. [2.14](#)
- [4] Debra L. Boutin. Identifying graph automorphisms using determining sets. *Electron. J. Combin.*, 13(1):R78, 2006. [4.9](#)
- [5] Irène Charon, Olivier Hudry, and Antoine Lobstein. Minimizing the size of an identifying or locating-dominating code in a graph is NP-hard. *Theoret. Comput. Sci.*, 290(3):2109–2120, 2003. [1](#)
- [6] Gary Chartrand, Michael Raines, and Ping Zhang. The directed distance dimension of oriented graphs. *Math. Bohem.*, 125(2):155–168, 2000. [4.2](#), [4.3](#)
- [7] N.J. Fine and H.S. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16(1):109–114, Feb., 1965. [2.13](#)
- [8] R. L. Graham, M. Grötschel, and L. Lovász, editors. *Handbook of combinatorics. Vol. 1, 2*. Elsevier Science B.V., Amsterdam; MIT Press, Cambridge, MA, 1995. [4.3](#)

- [9] Mark G. Karpovsky, Krishnendu Chakrabarty, and Lev B. Levitin. On a new class of codes for identifying vertices in graphs. *IEEE Trans. Inform. Theory*, 44(2):599–611, 1998. [1](#)
- [10] Samir Khuller, Balaji Raghavachari, and Azriel Rosenfeld. Landmarks in graphs. *Discrete Appl. Math.*, 70:217–229, 1996. [1](#)
- [11] Y. Kikuchi and Y. Shibata. On the domination numbers of generalized de Bruijn digraphs and generalized Kautz digraphs. *Inform. Process. Lett.*, 86:79–85, 2003. [4.2](#), [4.1](#)
- [12] Marilynn Livingston and Quentin F. Stout. Perfect dominating sets. *Congr. Numer.*, 79:187–203, 1990. [4.1](#)
- [13] Andréa Richa, Christian Scheideler, and Phillip Stevens. Self-stabilizing de Bruijn networks. In *Stabilization, Safety, and Security of Distributed Systems*, pages 416–430. Springer, 2011. [1](#)
- [14] Steven S. Skiena. *The Algorithm Design Manual*. Springer Publishing Company, Incorporated, 2nd edition, 2008. [1](#)